5. Summary
Entamoeba histolytica, an enteric protozoan parasite, is the causative agent of amoebic dysentry and amoebic liver abscess. It kills 100,000 people annually throughout the world and ranks third as a cause of death due to parasitic infection. Genome organization is so far not clearly understood in *E. histolytica*. The post genomic era has brought the realization that vast chunks of most genomes are composed of highly repetitive sequences—notably transposable elements (TEs). TEs are considered to be important genomic components in the same category as exons, promoters, enhancers, telomeres, etc. TEs by virtue of their transposition property have facilitated the shuffling and reorganization of the genome.

The genome of *E. histolytica* contains few DNA transposons, many non-LTR retrotransposons but no LTR retrotransposon. Three families of non-LTR retrotransposons called EhLINEs and their partners EhSINEs have been identified in *E. histolytica* genome. SINEs are non-autonomous non-LTR retrotransposons which utilize the transposition function encoded by partner LINE for their transposition. In many organisms including *E. histolytica* partner LINEs and SINEs share a stretch of common sequence at the 3'-end. EhLINE1 and EhSINE1 share a sequence of 74 nucleotides at their 3'-end. Although EhLINE1 is present in a few hundred copies per genome, it is not known whether a single active copy of this element exists in the *E. histolytica* genome. Genome sequence analysis showed that though most copies have multiple mutations, several entries in *E. histolytica* GSS database showed long complete ORFs. This raises the possibility that there may be at least a few functional copies of this retrotransposon in *E. histolytica*. Non-LTR retrotransposons encodes three recognized functions required for retrotransposition, namely reverse transcriptase (RT), endonuclease (EN), and nucleic acid binding protein (NB). The general model that has served to explain LINE retrotransposition is Target Primed Reverse Transcription. According to this model, the LINE encoded EN nicks the bottom strand of target site, generating 3'-OH group that primes reverse transcription of the element-RNA at the target site itself. The insertion specificity of the element is, therefore, guided to a large extent by target-site recognition property of the element-encoded EN. The non-LTR retrotransposons can be classified into two broad categories based on the nature of the endonuclease encoded by the
elements. One class encodes the apurinic endonuclease (APE), while the other encodes restriction enzyme-like endonuclease (Rel-ENDO). The EhLINE1 contains an Rel-ENDO domain downstream of a centrally located RT domain. However while most known members of Rel-ENDO containing elements insert in a sequence specific manner, this does not appear to be the case with EhLINE1. The element is found on all chromosomes in *E. histolytica*. The partner EhSINE1 is also widely dispersed throughout the genome.

The work carried out in this thesis mainly involved the study of biochemical properties of EN with respect to target site specificity of nicking *in vitro*. We have cloned and expressed the EN domain in *E. coli*. The purified protein could nick supercoiled plasmid (pBS) DNA to yield open circular and linear DNAs. The conserved motif PD..D was required for its activity. Genomic sequences flanking the sites of insertion of EhLINE1 and partner EhSINE1 showed both elements insert in non-sequence specific way. However a common feature was the presence of a short T-rich stretch just upstream of the target site duplication (TSD) in most insertion sites. To assay the activity of EN on its natural substrates we looked for pre-integration sites of EhLINE1/EhSINE1 in the *E. histolytica* genome. Only one such empty site for EhSINE1 was identified. When a 174 bp fragment containing the empty site was used as a substrate for EN, it showed some major nicking sites on the bottom strand. Of these, one prominent nick was at the exact site of insertion of EhSINE1. This showed that EhSINE1 could use the EhLINE1 encoded EN for its insertion. However, the pattern of nicks on the top strand was different from that of the bottom strand. The enzyme acted on many sites but none was as strongly preferred as seen in case of bottom strand. Sequence analysis of nicking sites revealed two important features- i) the most commonly nicked sequence on the bottom strand was 5'-AT-3' and ii) A C-residue (in top-strand sequence) was very frequently present 3 to 4 nucleotides downstream of the nick.

Further studies on substrate requirements revealed that a 27 bp (-11 to +16 with respect to the nick site) segment was sufficient as a substrate for EN. Analysis of nucleotide sequence preference showed that changing the nucleotides upstream of the nick influenced EN activity marginally whereas, changing the nucleotides downstream of
the nick had a much larger impact. However, the nucleotide sequences some distance away from the nicking site (both upstream and downstream) did not influence EN activity. Our result also showed that downstream C- residues (frequently found in most of the major nicking sites) did play a significant role in substrate recognition. From the mutation data the preferred recognition sequence for EN was deduced to be 5'-AATGC-3'. Several *E. histolytica* sequences were good substrates of the EN *in vitro* but the elements were not found inserted in many of these regions of the *E. histolytica* genome. Hence, other features of the target sites, including flanking sequences and target conformation, may be important.

We have also tested EN activity on a completely artificial substrate. Here also the enzyme made a prominent nick between A and T residues and increased nicking activity was observed when C nucleotides were placed immediately downstream of the nicking site. We have also studied the biochemical properties of the EhLINE1 EN protein. EN activity varied with temperature, pH as well as with salt concentration in the buffer. The enzyme activity was weak at high (500 mM) as well as at low (25 mM) NaCl concentration. Optimum activity was seen at 100 mM NaCl.

We have thus characterized the endonuclease domain encoded by EhLINE1 element. Although EhLINE1 encodes a restriction enzyme-like endonuclease and should insert at specific sites, it actually shows a very broad genomic distribution.

It will be extremely important and interesting for the future to understand how EhLINE1 encoded EN works *in vivo*, which will help us to understand how non-LTR retrotransposons shaped the *E. histolytica* genome during the course of evolution.